THE ASTROPHYSICAL JOURNAL, 580:L93–L96, 2002 December 1

STATISTICS OF NEUTRAL REGIONS DURING HYDROGEN REIONIZATION

ADI NUSSER,[1] ANDREW J. BENSON,[2] NAOSHI SUGIYAMA,[3,4] AND CEDRIC LACEY[5,6]

Received 2002 August 1; accepted 2002 October 18; published 2002 October 30

ABSTRACT

We present predictions for two statistical measures of the hydrogen reionization process at high redshift. The first statistic is the number of neutral segments identified in spectra of high-redshift QSOs as a function of their length. The second is the cross-correlation of neutral regions with possible sources of ionizing radiation. These independent probes are sensitive to the topology of the ionized regions. If reionization proceeded from high- to low-density regions, then the cross-correlation will be negative, while if voids were ionized first we expect a positive correlation and a relatively small number of long neutral segments. We test the sensitivity of these statistics for reionization by stars in high-redshift galaxies. The flux of ionizing radiation emitted from stars is estimated by identifying galaxies in an *N*-body simulation using a semianalytic galaxy formation model. The spatial distribution of ionized gas is traced in various models for the propagation of the ionization fronts. A model with ionization proceeding from high- to low-density regions is consistent with the observations of Becker et al., while models in which ionization begins in the lowest density regions appear to be inconsistent with the present data. Further data are required to confirm this result.

Subject headings: cosmology: observations — cosmology: theory — dark matter — intergalactic medium — large-scale structure of universe — quasars: absorption lines

1. INTRODUCTION

Early hydrogen reionization is a particularly interesting process in the high-redshift universe and is inevitably linked to the appearance of the first star-forming objects, at least those that served as sources of the ionizing radiation. If it occurred early enough, reionization imprints distinct features in maps of the cosmic microwave background (CMB) on arcminute angular scales (Vishniac 1987; Bruscoli et al. 2000; Benson et al. 2001, hereafter BNSL). Several aspects of hydrogen reionization remain uncertain despite the rapidly accumulating data on the high-redshift universe. For example, it is unclear what objects produce most of the ionizing radiation, although high-redshift galaxies are very strong candidates (Couchman & Rees 1986; Haiman & Loeb 1997; Ciardi et al. 2000). It is also unclear how the ionized regions develop in space (Miralda-Escudé, Haehnelt, & Rees 2000). The ionizing sources are likely to lie in high-density regions, but those regions do not necessarily ionize first; the ionizing photons may tunnel into less dense regions and ionize those first. Furthermore, the duration of reionization is unknown, and only a lower limit on the redshift marking the end of that epoch exists (Becker et al. 2001; Gunn & Peterson 1965).

In previous papers (BNSL; Liu et al. 2001), we examined how the CMB is affected by the reionization process and how future CMB maps can be used to extract information on that process. However, hydrogen reionization is currently best probed by spectra of high-redshift QSOs. Unlike maps of the CMB that are sensitive to line-of-sight integrals over the density and velocity of ionized gas, QSO spectra contain direct information on the local distribution of neutral hydrogen.

Recently, Becker et al. (2001) analyzed spectra of a sample of QSOs with redshifts between $z = 5.82$ and $z = 6.28$. In the spectrum of their highest redshift QSO ($z = 6.28$), the transmitted flux in the Lyα and Lyβ forest in the redshift stretch $5.95 < z < 6.16$ is consistent with zero, with a lower limit of 20 on the Lyα optical depth. This long stretch in redshift corresponds to a comoving distance of 60 h^{-1} Mpc in a universe with a cosmological constant of $\Lambda_0 = 0.7$ and matter density of $\Omega_0 = 0.3$. Becker et al. (2001) suggest that this long neutral region is a detection of the end of the hydrogen reionization era. To increase the signal-to-noise ratio, Becker et al. binned their spectra in 4 Å pixels. This prevented them from detecting small-scale dark windows that are likely to appear in the spectra as leftovers from the reionization epoch. A high-resolution spectrum for one of the quasars observed by Becker et al. was obtained by Djorgovski et al. (2001). The spectrum of this quasar ($z = 5.73$) was thoroughly analyzed by Djorgovski et al. (2001) and was found to contain several small dark windows signifying the detection of the trailing edge of the reionization epoch.

Motivated by the results of Becker et al. (2001) and Djorgovski et al. (2001), we examine here how the key ingredients in the reionization process can be probed by QSO spectra and future high-redshift galaxy surveys. Using the methodology of BNSL, we obtain predictions for two statistics. The first is the expected number of neutral segments longer than a given length, and the second is the cross-correlation function between the galaxies and neutral regions.

2. MODELING THE DEVELOPMENT OF IONIZED REGIONS

BNSL employed a semianalytical model for galaxy formation in a high-resolution *N*-body simulation of dark matter to estimate the amount of ionizing radiation produced by stars in high-redshift galaxies. Here we follow a similar procedure using the latest version of the GALFORM galaxy formation model (Cole et al. 2000) and the same ΛCDM simulation as BNSL (cf. Jenkins et al. 1998). This simulation has $\Omega_0 =$

[1] Physics Department and Space Science Institute, Technion-Israel Institute of Technology, Haifa 32000, Israel; adi@physics.technion.ac.il.
[2] California Institute of Technology, MC 105-24, 1200 East California Boulevard, Pasadena, CA 91125; abenson@astro.caltech.edu.
[3] Division of Theoretical Astrophysics, National Astronomical Observatory of Japan, Mitaka 181-8588, Japan; naoshi@yso.mtk.nao.ac.jp.
[4] Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Strasse 1, Postfach 1317, D-85741 Garching, Germany.
[5] International School for Advanced Studies, via Beirut 2–4, 34014 Trieste, Italy.
[6] Department of Physics, University of Durham, South Road, Durham DH1 3LE, UK.

0.3, a cosmological constant $\Lambda_0 = 0.7$, and a Hubble constant of $h = 0.7$ in units of 100 km s^{-1} Mpc^{-1}, and is normalized to produce the observed abundance of rich clusters at $z \approx 0$ according to Eke, Cole, & Frenk (1996). The simulation has a box of comoving length 141.3 h^{-1} Mpc and contains 256^3 dark matter particles.

Most of the ionizing photons emitted by stars are likely to be absorbed by gas and dust inside galaxies, and only a small fraction, $f_{\rm esc}$, escapes and becomes available for hydrogen ionization in the intergalactic medium (IGM; Leitherer et al. 1995). Assuming a value[7] for $f_{\rm esc}$, BNSL used the following models to follow the propagation of ionized regions in the simulation (see BSNL for details).

Model A (*growing front model*).—We ionize a spherical volume around each source (halo) with a radius equal to the ionization front radius for that halo assuming a large-scale uniform distribution of neutral hydrogen. Since the neutral hydrogen in the simulation is not uniformly distributed, and also because some spheres will overlap, the ionized volume will not contain the correct total mass of hydrogen. We therefore scale the radius of each sphere by a constant factor and keep repeating the procedure until the correct total mass has been ionized.

Model B (*high-density model*).—We simply rank the cells in the simulation volume by their density. We then completely ionize the gas in the densest cell. If this has not ionized enough hydrogen, we ionize the second densest cell. This process is repeated until the correct total mass of hydrogen has been ionized.

Model C (*low-density model*).—Same as model B, but we begin by ionizing the least dense cell, and work our way up to cells of greater and greater density (Miralda-Escudé et al. 2000).

Model D (*random spheres model*).—Same as model A, but the spheres are placed in the simulation entirely at random rather than on the dark matter halos.

Model E (*boundary model*).—We ionize a spherical region around each halo with a radius equal to the ionization front radius for that halo. This may ionize too much or not enough neutral hydrogen depending on the density of gas around each source. We therefore begin adding or removing cells at random from the boundaries of the already ionized regions until the required mass is ionized.

Models A, B, and E all are associated with ionization starting in the highest density regions. The differences among them will serve as an assessment of the sensitivity of the proposed statistics. Model C is motivated by Miralda-Escudé et al. (2000). In the random sphere model (D), the ionized bubbles are uncorrelated and serve to test the effect of the correlation. Model D also approximates a situation in which in some regions in space ionization proceeds from high densities while in others from the voids.

Guided by the observations of Becker et al. (2001), we will compute the number of neutral segments and the cross-correlations at three redshifts, $z = 6.67$, 6.22, and 5.80. Table 1 lists the volume filling factors (ratio of volume of ionized regions to total volume of the simulation box) in each of our five models for $f_{\rm esc} = 0.01$ (col. [3] in Table 1) and $f_{\rm esc} = 0.05$ (col. [4]). We will see later that the results of Becker et al. imply that the amount of ionizing radiation increases significantly between $z \approx 6.2$ and 5.8. Therefore, we also show results for a variable escape fraction, $f_{\rm esc}^{\rm var}$ (col. [5]), which equals 0.01 before $z = 6.22$ and increases linearly with time to 0.1 at $z = 5.8$. As expected, model C (low-density model) has the highest filling factor for a given $f_{\rm esc}$. For $f_{\rm esc} = 0.05$, the simulation box is fully ionized at $z = 5.8$ in all models.

[7] BNSL considered several models for the variation of $f_{\rm esc}$ from galaxy to galaxy. Here we adopt the simplest model, in which $f_{\rm esc}$ is constant for all galaxies.

TABLE 1
VOLUME FILLING FACTORS

Model (1)	z (2)	$f_{\rm esc} = 0.01$ (3)	$f_{\rm esc} = 0.05$ (4)	$f_{\rm esc}^{\rm var} = 0.1$ (5)
A	6.66	0.155	0.612	...
	6.22	0.187	0.892	...
	5.80	0.210	1	0.835
B	6.66	0.079	0.464	...
	6.22	0.101	0.688	...
	5.80	0.114	1	0.624
C	6.66	0.458	0.892	...
	6.22	0.536	1	...
	5.80	0.589	1	0.982
D	6.66	0.229	0.726	...
	6.22	0.270	0.919	...
	5.80	0.315	1	0.922
E	6.66	0.183	0.648	...
	6.22	0.217	0.937	...
	5.80	0.238	1	0.871

NOTE.—Here we show the volume filling factor of ionized regions in the simulation at $z = 6.66$, 6.22, and 5.80, for two constant values of $f_{\rm esc}$ (cols. [3] and [4]) and a variable fraction $f_{\rm esc}^{\rm var}$ (col. [5]) that increases linearly with time from 0.01 at $z = 6.22$ to 0.1 at $z = 5.80$.

3. THE STATISTICAL MEASURES

3.1. *The Number of Neutral Segments*

We are now in a position to compute the proposed statistics. We begin with $N(>L)$, the mean number of neutral (unionized) segments of length greater than L in a given redshift range in a line of sight (see Pentericci et al. 2002 and Barkana 2002 for similar statistical measures). We have the spatial distribution of ionized and neutral regions in the simulation as a function of redshift, for each of our five reionization scenarios (see Table 1). To compute $N(>L)$, we choose several random "lines of sight" in the output of the simulation at a given redshift. In each line of sight, we identify the neutral segments and tabulate their lengths, L, in comoving h^{-1} Mpc. A line of sight is obtained by starting from a grid point at the boundary of the simulation and going around the boundary of a rectangular slice of perimeter $4 \times 141\ h^{-1}$ Mpc (comoving) until we return to the starting point. (The shape of the path used to extract a line of sight makes no difference to our results. Using a rectangular path helps reduce the chance of pattern repetition.) This yields a total of 256 lines of sight, each spanning a redshift range corresponding to 564 h^{-1} Mpc (comoving). We then compute the mean $N(>L)$ from these lines of sight. For convenience, we normalize $N(>L)$ to a redshift span of 1 h^{-1} Gpc by multiplying the direct result obtained from the simulation by 1000/564. The $N(>L)$ (normalized to 1 h^{-1} Gpc) is shown in Figure 1 for $z = 6.66$ (*top*), 6.22 (*middle*), and 5.8 (*bottom*). The panels to the left show $N(>L)$ computed for $f_{\rm esc} = 0.01$ at these three redshifts. To the right, we show curves computed with $f_{\rm esc} = 0.05$ at $z = 6.66$ (*top*) and 6.22 (*middle*) and a variable $f_{\rm esc}^{\rm var}$ at $z = 5.8$ (*bottom*). We have also computed $N(>L)$ from lines of sight each of length 141 h^{-1} Mpc passing through the simulation box at random positions and found very similar results to those shown in the figure.

The cells of our computational grid are approximately

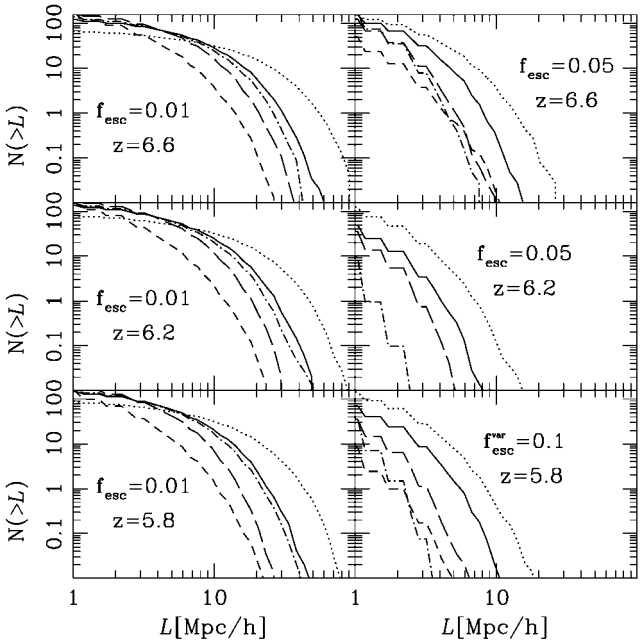

FIG. 1.—Mean number of neutral segments of length greater than L in lines of sight each of redshift span corresponding to 1 h^{-1} Gpc, as estimated from the simulation. The lines in each panel refer to the five ionization models: model A (*solid line*), B (*dotted line*), C (*short-dashed line*), D (*long-dashed line*), and E (*dash-dotted line*). There are no neutral regions in model C with $f_{esc} = 0.05$ at $z = 6.22$. Models with f_{esc}^{var} have $f_{esc} = 0.01$ for $z > 6.22$ and increase linearly to $f_{esc} = 0.1$ from $z = 6.22$ to $z = 5.80$.

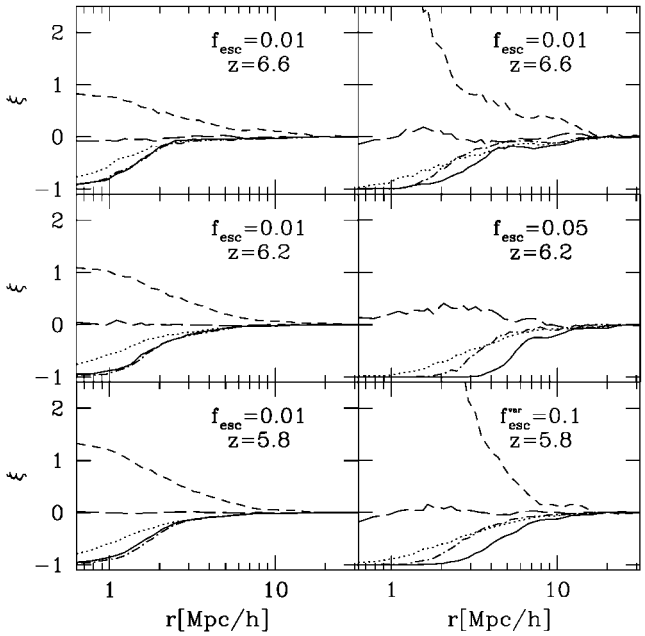

FIG. 2.—Cross-correlations between the galaxy distribution and neutral regions in the simulation. The notation of the lines is the same as in Fig. 1.

0.55 h^{-1} Mpc (comoving) in extent. According to BNSL, the characteristic smoothing length in the IGM at this redshift is approximately 0.2 h^{-1} Mpc (comoving). Thus, our current simulation does not fully resolve the structure of gas in the IGM. To assess the consequences of this limitation, we repeated our calculations using a lower resolution computational grid (128^3 cells instead of 256^3). We find that, at large L, the lengths of neutral regions are approximately 30% smaller when the higher resolution grid is used. Doubling the grid resolution to 512^3 cells (and therefore almost fully resolving the smoothing length) should make little difference to our results. We also checked that noise due to the finite number of particles is unimportant for the results presented here.

3.2. *The Cross-Correlation*

In addition to the distribution of neutral hydrogen, the cross-correlation, ξ, requires the spatial distribution of (potential) sources of the ionizing photons. In real observations, the distribution of neutral hydrogen is obtained from QSO spectra and the positions of sources from a high-redshift galaxy or QSO survey. Our goal here is simply to demonstrate that ξ can distinguish between various models for the propagation of the ionized regions. We will therefore simply compute ξ from the three-dimensional distribution of neutral hydrogen in the simulation; i.e., we will not address the question of how well ξ can be estimated from realistic mock observations where the neutral regions are identified in lines of sight.

The simulation box is divided into a 256^3 cubic grid. At each grid point $\boldsymbol{x}$, we define a quantity h to be zero if that point has been ionized and unity otherwise. We also use the cloud-in-cell method to derive the number density $n(\boldsymbol{x})$ of galaxies from the galaxy positions in the simulation. We include only those galaxies with ionizing luminosities sufficiently high to ensure the population is fully resolved in the N-body simulation. Because ionizing luminosity correlates only weakly with halo mass—since it depends so strongly on the star formation rate—this means we select only the most luminous sources—5.4, 10, and 11 × 10^{54} h^{-2} photons s^{-1} at $z = 5.8$, 6.2, and 6.7, respectively. These sources are rare and contribute only a small fraction to the total ionizing luminosity density of the universe (about 15% and 18% at $z = 6.22$ and 5.80, respectively). Denoting the average values of n and h by $\bar{n}$ and $\bar{h}$, respectively, we define the cross-correlation, ξ, as

$$\xi(\boldsymbol{r}) = \langle [n(\boldsymbol{x})/\bar{n} - 1][h(\boldsymbol{x} + \boldsymbol{r})/\bar{h} - 1] \rangle, \quad (1)$$

where the angle brackets imply averaging over all grid points $\boldsymbol{x}$. In practice, the calculation of ξ is done using the technique of fast Fourier transforms. Figure 2 demonstrates that ξ is sensitive to the ionization model. It is positive for model C (low density), almost vanishes for model D (random spheres), and is negative for models A, B, and E, which ionize dense regions first. However, ξ has a similar shape for all the high-density models (A, B, and E), and we expect that it will be difficult to distinguish between them in observational data. Nevertheless, they all are significantly different from either model C or D. So ξ should successfully discriminate among low-density, random ionization, and high-density models.

4. DISCUSSION

The statistic $N(>L)$, giving the number of neutral segments of length greater than L for a given total length of a QSO spectrum, is sensitive to the filling factor and the way in which ionization proceeds. The cross-correlation between candidate ionizing sources and neutral regions is less sensitive to the filling factor but is a more direct and robust probe of the propagation of the ionization fronts. In addition to QSO spectra, the cross-correlation requires a sample of candidates (galaxies and QSOs) for the ionizing radiation. Catalogs of galaxies and QSOs at high redshift are rapidly accumulating, making it pos-

sible to compute the QSO-flux and galaxy-flux correlations. Comparison between galaxy-flux and QSO-flux correlation functions will tell us whether galaxies or QSOs contributed most of the ionizing radiation.

Current observations do not allow a robust determination of $N(>L)$. The number of spectra needed to determine $N(>L)$ to within a given accuracy can be estimated by noting that the relative error on this function is $1/\sqrt{MN}$, where M is the number of observed QSO spectra covering the same redshift range. Nevertheless, we still can make general conclusions based on the Becker et al. (2001) result, assuming that the long Gunn-Peterson trough they observe is indeed a signature of reionization. Let us take the length of a spectrum in the Lyα forest at $z \approx 6$ to be $\sim 250\ h^{-1}$ Mpc, corresponding to the comoving distance between Lyα and Lyβ emission lines. An inspection of Figure 1 shows the following: (1) A completely neutral stretch of a comoving length of $60\ h^{-1}$ Mpc at $z \approx 6.2$ is inconsistent with a large filling factor. (2) The observations indicate that the chances of finding long neutral regions at $z < 5.94$ are tiny, while they are significant at higher redshift. This behavior seems inconsistent with our theoretical $N(>L)$ computed with constant $f_{\rm esc}$. If $f_{\rm esc} = 0.01$, then there are similar probabilities for finding long segments at $z = 6.22$ and $z = 5.80$. If $f_{\rm esc} = 0.05$, then the box is fully ionized at 5.80, while at $z = 6.22$ long segments are very rare. Therefore, the data favor models in which there is a significant increase in the amount of ionizing radiation in the IGM, due to either an increasing escape fraction over this redshift range or a much stronger evolution in the galaxy/QSO population than is predicted by our model. (3) Our model C, in which ionization proceeds from low- to high-density regions, seems to be inconsistent with an $\sim 60\ h^{-1}$ Mpc neutral region, even for escape fractions as low as $f_{\rm esc} = 0.01$.

We point out two caveats to the above discussion. First, model C assumes that reionization begins from the least dense regions. However, in cosmological simulations of radiative transfer (Gnedin 2000; Ciardi et al. 2000) reionization appears to begin from moderate density regions that are near the sources before proceeding to the lowest density regions. So model C (Miralda-Escudé et al. 2000) serves as only an approximation to the reionization process as seen in the simulations. The second caveat is that in observed spectra ionized regions with large optical depths can be confused with completely unionized regions and vice versa. By measuring Lyβ absorption, lower limits of about 20 on the Lyα optical depth can be obtained. Regions with large optical depth can arise in the presence of strong large-scale fluctuations in the ionizing background. To estimate the degree of the confusion, we have computed $N(>L)$ assuming that regions with optical depth larger than 20 are identified as unionized. We found that in this more detailed calculation $N(>L)$ is shifted to smaller L by less than a factor of 2 at large L. At small L, $N(>L)$ is reduced by a factor of less than 2, although this could likely be reduced with a more careful analysis, as suggested above. Nevertheless, the effect of fluctuations in the ionizing background on QSO spectra still needs to be quantified in detail.

This research is supported by the EC RTN network "The Physics of the Intergalactic Medium." A. N. is supported by the Israeli Academy of Science, the German Israeli Foundation for Scientific Research and Development, and the Technion V. P. R. Fund and Henri Gutwirth Promotion of Research. N. S. is supported by the Alexander von Humboldt Foundation and a Japanese Grant-in-Aid for Science Research Fund of the Ministry of Education, 14540290. C. G. L. is supported by PPARC. A. N. and A. J. B. wish to thank the National Astronomical Observatory of Japan in Mitaka for its hospitality and support.

REFERENCES

Barkana, R. 2002, NewA, 7, 85
Becker, R. H., et al. 2001, AJ, 122, 2850
Benson, A. J., Nusser, A., Sugiyama, N., & Lacey, C. G. 2001, MNRAS, 320, 153 (BNSL)
Bruscoli, M., Ferrara, A., Fabbri, R., & Ciardi, B. 2000, MNRAS, 318, 1068
Ciardi, B., Ferrara, A., Governato, F., & Jenkins, A. 2000, MNRAS, 314, 611
Cole, S., Lacey, C. G., Baugh, C. M., & Frenk, C. S. 2000, MNRAS, 319, 168
Couchman, H. M. P., & Rees, M. J. 1986, MNRAS, 221, 53
Djorgovski, S. G., Castro, S., Stern, D., & Mahabal, A. A. 2001, ApJ, 560, L5
Eke, V. R., Cole, S., & Frenk, C. S. 1996, MNRAS, 282, 263
Gnedin, N. 2000, ApJ, 535, 530
Gunn, J. E., & Peterson, B. A. 1965, ApJ, 142, 1633
Haiman, Z., & Loeb, A. 1997, ApJ, 483, 21
Jenkins, A., et al. 1998, ApJ, 499, 20
Leitherer, C., Ferguson, H., Heckman, T. M., & Lowenthal, J. D. 1995, ApJ, 454, L19
Liu, G. C., Sugiyama, N., Benson, A. J., Lacey, C. G., & Nusser, A. 2001, ApJ, 561, 504
Miralda-Escudé, J., Haehnelt, M., & Rees, M. J. 2000, ApJ, 530, 1
Pentericci, L., et al. 2002, AJ, 123, 2151
Vishniac, E. T. 1987, ApJ, 322, 597